



Floyd, J. (2020). Normative behaviourism as a solution to four problems in realism and non-ideal theory. *Critical Review of International Social and Political Philosophy*, 23(2), 137-162.  
<https://doi.org/10.1080/13698230.2018.1501538>

Peer reviewed version

Link to published version (if available):  
[10.1080/13698230.2018.1501538](https://doi.org/10.1080/13698230.2018.1501538)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Taylor & Francis at <https://www.tandfonline.com/doi/abs/10.1080/13698230.2018.1501538> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Normative behaviourism as a solution to four problems in realism and non-ideal theory

[PENULTIMATE DRAFT – PLEASE SEE PUBLISHED VERSION IN ORDER TO CITE, available at: <https://www.tandfonline.com/doi/abs/10.1080/13698230.2018.1501538> ]

## Abstract

This article advances the case for ‘normative behaviourism’ – a new way of doing political philosophy that tries to turn facts about observable patterns of behaviour, as produced by different political systems, into grounds for specific political principles. This approach is applied to four distinct problems at the heart of the ideal/nonideal theory and moralism/realism debates: (1) How to distinguish good from bad idealisations; (2) how to rank options of variable feasibility, cost, and danger; (3) how to distinguish legitimate acceptance of a given political system from acceptance based on coercion or false consciousness; (4) how to translate abstract principles into concrete institutions. Objections against the general viability of normative behaviourism, and against the types of behaviour it tracks, are also considered.

## Keywords

Normative behaviourism; ideal theory; non-ideal theory; moralism; realism; methodology; facts and principles; Cohen; justification.

## 1.1 Introduction

In this article I discuss a new method of justifying political principles that either solves or alleviates four of the most pressing problems at the heart of two distinct though closely related debates: ideal vs. nonideal theory and moralism vs. realism. This method, *normative behaviourism*, treats facts, and in particular facts about observable patterns of behaviour, as the basic materials out of which political principles are justified<sup>1</sup> (Floyd, 2011a; Floyd, 2016a; Floyd, 2017b). The four problems I have in mind are these:

- (1) *The idealisation problem* – how should we distinguish good idealisations from bad ones in the justification of political principles?
- (2) *The ranking problem* – how should we rank different political options in the real world, given their variable feasibility, cost, and danger?
- (3) *The legitimacy problem* - how should we distinguish legitimate acceptance of a given political system from acceptance based on coercion or false consciousness?
- (4) *The institutional problem* – how should we translate political principles into viable political institutions?

My argument moves through three stages. First, I provide a sketch of normative behaviourism. Second, I explain how it solves or helps to solve the four problems described. Third, I defend it against objections to do with (1) its general viability and (2) the particular kinds of behaviour it focuses on.

## 1.2 Normative behaviourism

Consider first here that the standard way of doing political philosophy is to try and ground political principles in normative *thoughts*, by which I mean thoughts about what should and should not be the case within the world, and what we should and should not do within it<sup>2</sup>. These thoughts come in three forms<sup>3</sup> (Floyd, 2017b, 120):

- (1) *Impartial choices of political system*: Decisions made in hypothetical situations in which the chooser of a political system is subject to various constraints of (a) knowledge (what they know of themselves and the world) and (b) consequence (what that choice will mean for them), e.g. ‘the original position’.
- (2) *Considered judgements*: Reflected-upon verdicts given by informed and intelligent people about different economic, legal, political, and social procedures and states of affairs, e.g. ‘slavery is wrong’.
- (3) *Intuitive choices of abstract principle*: Non-inferential gut-responses to a choice between two or more hypothetical actions, each of which represents a different principle, e.g. ‘the trolley dilemma’.

The hope behind these thoughts is that they can be treated as more or less direct expressions of political principles. We would then be *bound* by such principles, subject to certain caveats<sup>4</sup>, just by *having* such thoughts; that is, just by producing the kinds of choice, judgement, and intuition just described. I call this standard model of justifying political principles *mentalism* (Floyd, 2017b, 99-165). When political philosophers claim that a set of political principles X violates our intuition P, or that Y is an interpretation of our considered judgment Q, or that Z is just that set which we would choose in a situation defined and justified by, in tandem, both P and Q, this is the model they are using.

Note that is also the model used by *moral* philosophers when generating principles for *individuals* (call them *moral principles*), which is important, given that *political principles* are often derived from these, and thus only indirectly from the relevant normative thought(s). Political principles can therefore be *expressed* more or less directly by those thoughts. Consider, for example, Singer’s argument that because we *would* save a boy drowning in a pond if we could do so at little cost to ourselves, we *should* give money to poor individuals on the other side of the world (Singer, 1972). Here we move from a particular normative thought (the intuitive abstract choice to save the child), to a particular moral principle (help others in great need when you can do so at little cost to yourself). But we could also move one step further and say that, because we subscribe to *that* moral principle, we should also subscribe to a further *political* principle (rich countries should help poorer ones when they could do so at little cost to themselves). We would thus have moved from the initial thought to the eventual political principle, regardless of whether or not we think the latter was, properly speaking, already expressed by the former. Obviously we can only do so if we take the three situations (rescue, individual aid, and state aid) to be normatively analogous, but that problem is always there. If no two situations are normatively analogous, then there are no principles of any kind, given the kind of multiple-case applicability required for a principle to truly *be* a principle.

With this standard model in mind, we can now make sense of normative behaviourism. In contrast to mentalism, which relies upon three types of *thought*, normative behaviourism relies upon two types of *action*: (1) insurrection, and (2) crime, each of which is treated as a

*rejection* of the political principles currently governing the individual undertaking the action. On this way of doing things, actions function as verdicts on the political system that produces them. For example, consider how one form of mentalism – constructivism – tries to ground political principles in the hypothetical choices we imagine we would make in various imaginary situations. Normative behaviourism, in parallel, tries to ground political principles in the choices we *do* make in a variety of real political situations. It does this by treating certain kinds of behaviour as being expressive of certain kinds of choices; by treating those choices as expressive of certain political preferences; and by treating those preferences as expressive of, and thus grounds for, political principles. So, whereas mentalism says that because people *would* choose A, they are committed to principle B, normative behaviourism says that because people *do* behave in manner X, they are committed to principle Y.

A key idea here, clearly, is the idea of *expression*. I said earlier that in mentalism we treat things like our considered judgements as being expressive of particular moral or political principles, with the task for moral and political philosophers then being to work out just what those principles are. Normative behaviourism works similarly, given that it treats certain forms of action as being expressive of principles, just insofar as those actions express preferences either for or against the political principles that define the political environment in which those actions take place (with that for-or-against quality being the thing that makes them, ultimately, a kind of *verdict*). Or, more precisely, we treat institutions and policies, at one level, as expressions of political principles, and then, on another level, actions taken in response to those institutions and policies as expressions of verdicts on those principles. For example, authoritarian institutions are an expression of authoritarian principles, with action taken against those institutions, in turn, an expression of a verdict on those principles. As a result, we might say that the obvious slogan for normative behaviourism would be ‘actions speak louder than words’. But that’s not quite right. As will become clear, what really matters is that certain kinds of action speak with both more *clarity* and more *commitment* than certain kinds of *thought*.

These kinds of action, as noted, are insurrection and crime, but why these two? The reason is that these two forms of behaviour have a common quality that is crucial to understanding the kind of clarity and commitment just noted. This quality is the level of personal *risk* that insurrection and crime tend to involve, and thus the fact that people tend only to partake of them when they find their current lives unbearable, due to a lack of various goods. Such ‘goods’ might include a minimal income, access to gainful employment, a particular set of rights, and a certain kind of political regime. But only *might*, because according to normative behaviourism – and this is another key point – rather than deducing those things in advance, as mentalism would, by interpreting them as entailments of a set of political principles that is in turn entailed by patterns in our prior normative thoughts, we induce them by revealing the causal patterns behind the two types of behaviour described, and in particular the political system, and thus set of political principles, that produces a minimum of both. It is thus the behaviour that defines the goods, not the goods that define the behaviour.

With these ideas in mind then, precisely *which* behavioural patterns, in the real world, does normative behaviourism try to turn into political principles? The kind of patterns it works with are the kind produced by, amongst other things, the *Polity*, *ACLP*, and *Freedom House* datasets<sup>5</sup> – the thrust of which is that *liberal-democracies* produce *less insurrection* than any other kind of political system in recorded history. In addition, and drawing on the second type of action in which we’re interested here – crime – it also draws on recent criminological

research showing that *more egalitarian* liberal democracies generate *less crime* (Kelly, 2000; Neumayer, 2005; Wilkinson & Pickett, 2009). Normative behaviourism therefore offers up the possibility of something very interesting for contemporary political philosophy: a *pseudo-empirical justification for the defining principles of egalitarian liberal democracy*. This is because, whereas conventional (mentalist) political philosophy asks such questions as ‘what do our intuitions about fairness tell us about social justice?’, or ‘how would we respond to different political choices in various abstract and hypothetical situations?’, normative behaviourism asks ‘how do we respond, in terms of specific behaviours, to different political systems in practice?’. According to normative behaviourism, it is these real responses, observed in the real world, that ultimately give us the measure of a particular political system, and in turn the political principles it expresses<sup>6</sup>.

At this point, however, some caveats are needed. First, we might doubt normative behaviourism’s helpfulness, despite what has been said so far, unless we can see how it might help us with various problems in the literature with which we are already wrestling. Second, we might doubt its general viability unless we can see how it might rebut certain objections, such as the claim that it can only justify its choice of behavioural measures by falling back on just the kind of mentalism from which it was distinguished above. I turn to the first of these issues in the next section and to the second in the final part of the article.

### 1.3 The idealisation problem

The first problem to consider here concerns idealisations, which come in two forms in political philosophy. First, we idealise *the world* when we imagine, say, what it would be like if everybody had equal resources, or if just a few of us were stranded on a desert island. Second, we idealise *human beings* when we imagine, say, what they would be like if free of prejudice, ignorance, or conceptions of the good. We do this in order to identify and, simultaneously, justify political principles. For example, by asking what principles an idealised person would choose, we not only reveal new principles to ourselves, but also give ourselves a reason to adopt them, just because they are what such a person would choose. Similarly, by asking what principles we would choose in, say, a context of abundant resources in which everyone agreed to comply with the correct principles of justice, we hope to capture ideals of a sort that we could aim at right now.

The challenge is to distinguish good from bad idealisations. In general, we might say that *good idealisations* remove morally irrelevant or politically/economically/socially unnecessary facts about our existence in order to isolate the political principles that should guide our attempts to improve it, whereas *bad idealisations* render those principles inapplicable in real-life circumstances, either through not offering guidance, given the gap between simplified-ideal and complicated-real circumstances - the guidance critique (Valentini, 2009) - or through being dangerous, on account of their tendency to conceal, and thus legitimise, features that those principles ought to address, such as patriarchy or racial discrimination - the ideology critique (Mills, 2005)<sup>7</sup>. How though do we tell the one from the other? And notice, although if we incorporate too much idealisation our principles will be either useless or dangerous, there is also danger the other way. Incorporate too little and you end up accepting *too much* of the existing status quo. Now you run the risk not just of being unable to guide improvement of the world (for you have already accepted it) but also of legitimating the kinds of features that you would want your principles to reject. The guidance and ideology critiques, in other words, and this is under-appreciated, apply in both directions.

Normative behaviourism avoids this problem by avoiding idealisations altogether. Rather than contemplating imaginary people and imaginary worlds in order to generate ideal principles, and then worrying about *how* those principles apply to reality, normative behaviourism begins, as realists insist we should, with facts about how things already are (Geuss, 2009, 9). Unlike realism, however, it does not do this in order to more severely *constrain* our ideal principles, by saying that the political world is much more *difficult* to improve than political philosophers generally assume, and then referring to the dictum *ought-implies-can*. It relies instead on *is-to-ought*, by moving from the ‘is’ of how people behave under different institutions, and thus in response to the principles those institutions represent, to an ‘ought’ of what principles we ought to adopt.

This move, however, quite apart from the question of general is-to-ought viability discussed later on, also encounters what I call here the ‘punishing moral saints’ problem. To see this problem, consider an example commonly used to indict ideal theory - its alleged ignoring of our dependence, not just on the *care* of other human beings, but in particular on the care of *women*. The standard accusation is that if one idealises human beings as unattached and autonomous creatures, and then asks what principles such creatures would choose, one will ignore (and effectively legitimate) both the reality of this care and the difficulties generated for women in a world that ignores it (Mills, 2005, 178)<sup>8</sup>. Normative behaviourism, in turn, struggles with this problem in its own unique way: by struggling with the objection that if women commit neither crimes nor acts of insurrection under patriarchal political systems (or at least systems which permit patriarchal social arrangements) then both they and their unacknowledged burden will be ignored and, again, legitimated.

There are several possible responses to this problem. One is that the situation of these women is *analogous* to other situations (slavery, feudalism, etc.) in which we know people turn to insurrection and crime, and is for that reason indicted by normative behaviourism. A second is that normative behaviourism is always open to new political experiments (it only insists that those experiments remain unjustified until enough people have experienced them in practice), including experiments in which the situation of these women is improved, and is for that reason not *bound* to the principles operating in that situation, especially if it turns out that the relevant behavioural patterns improve in the wake of the that improvement<sup>9</sup>. A third is that, although an egalitarian liberal democracy might permit the kind of situation described, it also *contains within it* the elements of the necessary redress, as a result of its democratic institutions. A fourth is that, even if women and their advocates do not use those institutions themselves, their mere presence *encourages* the kind of insurrectionary behaviour associated most famously with the suffragettes (vandalism, arson, assault, etc.). This matters because, according to normative behaviourism, even a small decrease in insurrection makes one set of principles superior to another, meaning that a democracy lacking such insurrection, perhaps because it had already responded to it, would be better than one that contained it.

The third and fourth responses are the key ones here. The problem with the first is that we might struggle to pin down the required analogousness. The problem with the second is that, if we were restricted to it, normative behaviourism would be depressingly uncritical, because even if it is the only route to the *truth* (assuming, as normative behaviourism argues, that political principles can only be *fully justified* by their record in power), it would offer considerably less guidance than we would like in the political present. The third and fourth responses, by contrast, avoid these problems. The third does so by relying on a key causal

ingredient of democracy's stability: its malleability under pressure, and thus capacity for progress without collapse. This ingredient not only explains the peaceful approach to protest taken by the *suffragists*, but also illustrates that, if we understand the full causal story behind the normative behaviourist success of a given set of a political principles, we can often eliminate worries of the kind under discussion<sup>10</sup>. The fourth response then becomes a justificatory bonus, and especially so given the parallel between normative behaviourism and the thinking of the actors involved. Consider here that whereas the normative behaviourist says 'actions, not thoughts', the law-breaking *suffragettes* said 'deeds, not words' (Pankhurst, 1979, 38). So, normative behaviourism avoids the usual problems of idealisations by avoiding them altogether, yet does so without legitimating problematic features of our world that we might not have noticed, given its recommendation of principles which, in practice, afford great scope for progress.

#### 1.4 The ranking problem

Now to our second problem. Consider here first that even if we *knew* both that justice is the only political value and what principles of justice would apply in the best possible world, we still have to decide, in the present: (1) whether we correct the *greatest* contemporary injustices - the *comparative* approach (Sen, 2009) - or adopt the *quickest* route to perfect justice, even if that meant increasing injustice in the meanwhile - the *transitional* approach (Simmons, 2010); (2) how we should compare variably *feasible* attempts to correct injustice (Gilabert & Lawford-Smith, 2012; Lawford-Smith, 2012; Hamlin & Stemplowska, 2012); (3) how we should compare variably *expensive* attempts (Floyd, 2011b); and (4) how we should compare variably *dangerous* attempts (Schmitz, 2011, 775). Consider second that if we do *not* know these things, the problem is even harder. If, as realists often insist, different political values are both incommensurable and incompatible when fully realised, it might follow either that no rational guidance is available for such ranking, or (especially if politics is as they describe) that we are restricted to a very minimal satisficing ambition, such as order or peace<sup>11</sup>.

These problems explain why there is so much acceptance of the place of intuition and judgement in the literature. Even when our ideal principles are lexically *ordered* (as they are by Rawls), and even when knowing the justificatory *reasons* behind those principles helps us with the necessary trade-offs (Swift, 2008, 365), there is still widespread acceptance that the difficulties described will ultimately only be resolved by the instinctive balancing of either theorists, politicians, or democratic populations (Floyd, 2010). All of which leaves us with the following problem: different theorists, politicians, and populations choose *different balances*.

Normative behaviourism eases these problems the same way utilitarianism does: by providing a *metric* for the ranking of political options. In short, options generating less discontent (as expressed by insurrection and crime) are judged better than options generating more. So, as regards the transitional/comparative trade-off, normative behaviourism notes that individuals *already* respond to political arrangements in terms of such concerns. As a result, better trade-offs are ones that generate less expressed discontent. So, problem solved? Unfortunately not: The problem is that even if one grants the normative relevance of this discontent, particular compromises are normally both too complex and too unimportant (as judged by those who have to live with them) to be accurately tracked by the relatively crude criteria of insurrection and crime. As a result, normative behaviourism will normally do

better to evaluate our *procedures* for striking such compromises than the compromises themselves. Better political systems, on this view, are ones that tend to strike better compromises between the transitional and comparative concerns described, across all values, and as judged by the people who have to live with them<sup>12</sup>. So, although normative behaviourism only sometimes (if historical precedents apply) identifies the best compromise on a given issue, it will always be able to identify those political procedures which, in general, have been deemed better at making such compromises.

This procedural approach is crucial given that most political options involve a balance, not just of transitional/comparative concerns for a given value, or even of transitional/comparative concerns for multiple values, but also, as noted, of feasibility, expensiveness, and risk. Normative behaviourism therefore solves the ranking problem, for the most part, not by telling us how to rank every set of options, but again by recommending a particular ranking procedure, based on how different procedures (i.e. different political systems) have performed in the past, in terms of their behavioural consequences. Yet this is only the most part. As noted, there will be exceptional cases in which historical precedent guides us with a particular set, and these come in two forms. First, when the options themselves are changes to the political system. In this case normative behaviourism's procedural guidance is also option-selecting, given that it recommends whatever political system generates the least discontent. Second, when our options are both sufficiently important and historically prefigured. In this case, because we know that different options produce various levels of discontent, we rank them according to those variations. Normative behaviourism, therefore, solves the ranking problem by either providing historical precursors for our current options, or by recommending procedures by which we should select amongst them.

### **1.5 The legitimacy problem**

Note that when responding to both the idealisation and ranking problems, normative behaviourism is part of a wider taste for process over substance, though the usual claim is to say that there are *legitimate* procedures we could adopt (Mason, 2010). In ideal-theory mode, for example, we might prescribe deliberative or majoritarian democratic procedures to solve various problems<sup>13</sup>. In realist mode, we might say that what counts as legitimate is simply whatever is accepted as legitimate by the population in question (Williams, 2005; Horton, 2010; Newey, 2010). Both routes, however, encounter problems. The first, due to the grounding such procedures require, struggles to escape the kind of value-conflict it wants to transcend. The second, in turn, struggles to distinguish authentic acceptance from acceptance based on either coercion or false consciousness. This then encourages attempts to distinguish between reasonable and unreasonable acceptance, and ultimately between what a reasonable and an unreasonable person could accept, which means, in parallel with the ideal response, that rather than rising above what appears to be intractable value-conflict, realists can fall right back into it<sup>14</sup>.

Normative behaviourism, as noted, shares the general procedural strategy described, but also the realist tactic of resting authentic legitimacy on perceived legitimacy. As a result, it has to distinguish, as realism does, between genuine acceptance and ideological or coerced acceptance. But how? The key is to judge political principles, not on what cultural spokesmen say about them, or on what the majority say about them, or even on what the majority would think about them if asked to choose between their system and others under hypothetical



conditions, but on how people *behave* under different systems in history. In other words, if more people find themselves turning to insurrection and crime under a given regime than under alternative feasible regimes, then that regime is indicted, even if *at present* the majority declares it legitimate. They might, after all, only consider it legitimate because they do not know what life would be like under alternative systems, or believe their current leader is the son of god, or fear the persecution that would follow if they spoke out – all factors which explain both the temporary success of a system *and* its long-term failure. Remember here that, according to normative behaviourism, actions speak louder than either words or, ultimately, thoughts. For example, you might *think* libertarianism fair despite *committing* repeated crimes under a system that expressed it, just because the life it facilitates for you is so awful. So, although you might consider your present political system absolutely fine, what matters is whether, say, hundreds of millions of people have rebelled against it in the past. If its historical record is poor, then its political present - perhaps buoyed by natural-resource money, or steep but unsustainable economic growth, or a charismatic leader whose star will eventually wane - is comparatively irrelevant.

We can expand this point by responding both to the realist claim that we should begin with the ‘facts’ about how people really are, and to the problem that it is not really clear just which facts count. Facts about rape, prejudice, and abuse? Facts about voting habits? Facts about economic choices? Again, according to normative behaviourism, the facts that count when assessing political principles (and thus their relative legitimacy when in power) are facts about behavioural trends that both *vary according to variations in the governing political principles*, and which can be treated as *expressions of discontent*, either with the political regime directly (in the case of insurrection) or with the kinds of life it facilitates (in the case of crime). Yet this reveals two problems for normative behaviourism. First, an *information* problem - authoritarian countries hide statistics about crime the way they hide statistics about suicide. Second, a *scalar* problem – how many people have to turn to crime or insurrection before a set of principles is indicted? This is why normative behaviourism adopts a *comparative* approach to the assessment of sets of principles. It aims for the system that produces the *least* crime and insurrection, not the system that produces *no* behaviour of either kind. Clearly, people turn to things like theft and terrorism under all systems; our aim is simply to *minimise* the motivation behind them. Relatedly, it looks for *large* numbers and *long-term* trends. This means that we cannot *justify* a new experimented-with set of principles on the basis of a few years of success in a single context. We can however, at least sometimes, *reject* such experiments, by noting that all the key elements of an earlier failed system are present within it. In these cases we know that the ‘new’ experiment will *ultimately* fail, at least once temporary environmental conditions have been removed. Consider, for example, that although apartheid failed *because* it produced too much discontent, it failed *when* it failed because the Cold War ended.

## 1.6 The institutional problem

We come now to our fourth problem, which starts with a complaint from both realists and non-ideal theorists that political philosophy focuses on *principles* at the expense of *institutions*<sup>15</sup>. This complaint exposes two difficulties: (1) *under-determination* – if abstract principles are too indeterminate to generate specific institutional prescriptions, then they fail to offer guidance; and (2) *feasibility* – if those principles entail unfeasible institutions, and *ought-implies-can*, then this might invalidate them altogether, assuming that *cannot-implies-ought-not*. Of course, we might respond to that second challenge by saying that impossible

ideals can still guide us, if only by providing a measure of the second-best (Jubb, 2012), but we discussed that earlier. As noted with the ranking problem, if different possibilities deviate from the ideal in different ways, what counts as second best? The key point here, by contrast, is that even if one *knows* what kind of outcome or principle should be aimed at, we still need to know what that means in terms of *institutional prescriptions*, given that no outcome or principle can be realised *without* the right institutions. Maximising happiness, ensuring security, guaranteeing rights – all these things require institutions which, in practice, could either promote or undermine our ambitions.

Normative behaviourism addresses this problem by moving from *proven* institutions to abstract principles (is-to-ought), and thus recommending, *from the start*, feasible institutional prescriptions (thus avoiding cannot-implies-ought-not). That is, it begins with the particular set of institutions that generates the least insurrection and crime in practice, and then says that whatever principles those institutions express are, for that reason, justified. As a result, it seems to solve the two problems just described, *under-determination* and *feasibility*. Or does it? A clause that needs adding to the conclusion of the previous section is that a set of institutions is legitimate, not just when supported by the historical-behavioural record, but also *when it is the best possible in the circumstances*. Yet that clause creates its own challenge. When we lack the resources required to achieve, overnight, the particular set of institutions identified as ‘best’ by normative behaviourism, what should we do instead? In such circumstances we might conclude that, because the best is *infeasible*, normative behaviourism is too *indeterminate* to guide us. In other words, even if the best is possible in the West, normative behaviourism might be useless for the rest.

Strictly speaking, however, there is no *evidence* that introducing egalitarian-liberal-democracy overnight is impossible. History is full of societies which, after standing still for centuries, transformed their economic and political arrangements at relative speed. Think, for example, of some of the consolidated democracies of East Asia. The problem is that doing so requires *more* than just institutional design. It takes power, a lack of external interference, and a succession of wise leaders capable of persuading and unifying the people with inspiring rhetoric, all of which are naturally beyond the reach of any form of political *philosophy*. As a result, we could then leave things there, and say that even if there are never conditions in which, say, crony-capitalist dictatorship is *possible* and democracy *impossible*, we do often lack those further things required to convert possibility into reality. Of course the normative behaviourist, like all political philosophers, lacks the power to enforce her advice, but how could it be otherwise? In this respect she is the doctor who cannot *make* the patient take their medicine - it doesn’t mean she’s given the *wrong* medicine. We might even then say that, although normative behaviourism only deals in feasible and precise institutional prescriptions, it is nonetheless *transcendental* (in Sen’s terms), given that it focuses on the best, and *transitional* (in Simmons’ terms), given that it tells us to head for the best by the quickest route available.

We should, however, at least partially resist that conclusion. Although it certainly helps with this problem to give precise and feasible prescriptions which, given the right background conditions, could readily be put into practice, it might be that we can still do more. At the least, we can say that, until those conditions emerge, *more* legitimate institutions are ones that serve to *reduce* expressed discontent. This claim dovetails with a point made in response to both the idealisation and legitimacy problems: that we cautiously *experiment* with new policies and institutional amendments when we do not know their consequences, but only

*justify* any given effort, after the event, by showing that it has reduced overall levels of expressed discontent.

### 1.7 Normative behaviourism defended, pt. 1: the paradox/fallacy objection

In response to all that has been said so far, several objections could be made against the general viability of normative behaviourism, about which more has been said elsewhere (Floyd, 2017b, 166-254). For example, one might think (1) that normative behaviourism is excessively conservative, given that it apparently only justifies the historically-tested status quo; (2) that it cannot account for the collective action problems and unintended consequences that mediate between individual actions and political outcomes; and (3) that it ends up ‘justifying’ just about everything human beings do. One objection, however, is more important than all of these, and needs to be addressed here in order to give the preceding arguments their relevance. This objection says that normative behaviourism must be either paradoxical or violate the naturalistic fallacy. It would be paradoxical if (1) it claimed political principles should be grounded in actions rather than thoughts, and (2) in order to explain the normative relevance of those actions, it relied upon principles of a sort that could *only* be justified by the kinds of thoughts rejected under (1). It would violate the naturalistic fallacy if it tried to move from an ‘is’ of behaviour to an ‘ought’ of political principles *without* such thought-justified-principles.

I reject this objection here by rejecting the idea that *facts cannot ground principles*. Consider Cohen’s well-known argument that a fact cannot ground a principle in the absence of a prior principle explaining the relevance of that fact (Cohen, 2008). For example, for the fact ‘keeping promises helps human beings pursue their long-term projects’ to ground the principle ‘we should keep our promises’, there must be some prior principle of the form ‘we should help people to pursue their projects’ (*Ibid*, 234). In what follows, I offer two arguments against this claim, the first of which rejects its *relevance* and the second its *truth*. If either argument succeeds, the objection is rebutted.

First, *relevance*. Even if Cohen is right about the ultimate dependence of facts on principles, facts matter *more* than *ultimate principles* when justifying either *moral* or *political principles*. This is because ultimate principles are so *abstract*, due to their fact-independence, as to be virtually useless, normatively speaking. Consider again his ‘promises and projects’ example. Even if he is right that the normative relevance of the offered fact depends upon his suggested prior principle, there remains a question about the grounding of *that* principle. If we ask ‘why should we help people with their projects?’, surely we would give an answer of the form ‘because their projects are crucial to their well-being’. If so, Cohen’s second principle would *itself* be grounded in a second *fact* about human nature.

But that claim is not news to Cohen. He is perfectly comfortable with the idea of a fact grounding a principle. All he wants to deny is that they can do so in the *absence* of a *further* principle capable of explaining the *relevance* of that fact. He can therefore happily respond to my point about fact-dependence by saying, for example, that that fact is itself only relevant (in the sense of normatively grounding the principle above it) because it depends on a *third* principle of the form ‘we should promote human well-being’. As a result, we would then be returned, once more, to a position of *principle-dependence*.

Yet how far can this reasoning run? In order to find out, we need to bear in mind four points. First, that principles can be more or less *general* – ‘don’t lie’, ‘don’t lie to your friends’, and ‘don’t lie to your best friend’ are all principles. Second, that when we say X *grounds* Y, we do not normally mean that X functions as *ultimate* grounds for Y, given that X itself might need further grounding. For example, if we say that utilitarianism grounds the harm principle, we do not necessarily mean, and are unlikely to mean, that utilitarianism has or needs no further grounds. Third, that for one principle to ground another, it need not be the case that both principles are of the *same kind*. For example, when we say that utilitarianism grounds the harm principle, we may well be (a) grounding a *political* principle ‘governments should only use their coercive power in order to prevent individuals from harming one another’ in a *moral* principle of the form ‘maximise happiness in the world’, and (b) grounding *that* principle in a *methodological* principle of the form ‘adopt whatever set of moral principles is expressed by your intuitive responses to different moral dilemmas’. Fourth, that all of the principles considered so far are *fact-dependent*.

This last point is the key one. Consider here something that is easily overlooked: that the move from utilitarianism to the harm principle requires human-behavioural-facts about the consequences of adopting that principle, just as the move from the described methodological principle to utilitarianism requires facts about *what our intuitions are*. As a result, it appears to be a *truth*, or so I would suggest, that *thoughts*, like *actions*, are *facts*. For example, if I *intuitively think* that I should save a drowning child when I could do so at little cost to myself, then it is a *fact* that I have that intuition. This means at least two things: First, that mentalism (and thus conventional political philosophy) is just as dependent on facts as normative behaviourism, given that the one relies on facts about *thoughts* and the other on facts about *actions*. Second, we could now quite easily talk of the *mentalistic fallacy*, if it is true not just (1) that conventional political philosophy relies upon a certain class of facts, but also (2) that it is therefore confronted with the naturalistic fallacy, as it tries to move from mental ‘is’ to political ‘ought’, when it had thought itself exempt from such confrontation.

But there is also something else here. There is the truth that, even if we can find a further, fact-free ultimate principle capable of grounding the described methodological principle (and there might be several fact-free principles on the way to a fact-free *ultimate* principle), it will still be the case that, as far as political philosophy is concerned, fact-dependent principles are *where the action is*. This is because, in political philosophy, if one can show, for example (1) that our normative thoughts are contractarian in character, (2) that they require, politically, a reasonable rejection test of different political principles, and (3) that only one set of principles passes that test, then it will generally be considered that one has provided *all the grounding one needs*. In fact, even in the most fundamental and far-reaching works of *moral* philosophy, the standard problem explored is that of understanding what it *means* to be moral, or what morality *requires* (Rawls, 1999, 286-302; Scanlon, 1998, 148). The question, that is, of why one *should* be moral, or of what *motivation* is required for morality, is generally put to one side (normally because it is deemed either unanswerable or uninteresting). But then if that is acceptable, so is normative behaviourism. Assuming that the *action-expressed* political-principle-preferences tracked by normative behaviourism (i.e. insurrection and crime) are somehow equivalent to the *thought-expressed* political-principle-preferences studied by conventional political philosophy (e.g. considered judgements), then there is no reason why normative behaviourism cannot share the same starting point. Admittedly, I have not demonstrated this ‘equivalence’, or disproved the claim that normative behaviourism depends upon prior methodological principles that are themselves thought-

dependent, but that is not the point. The point is that Cohen's argument is now *irrelevant*, given that, even if facts cannot go all the way down, they still go far enough to reach the point where political philosophy normally begins.

Now we turn to the second argument - against the *truth* of Cohen's claim. Consider then this problem of finding grounds for the *methodological* principle just considered (which was, to reiterate: 'adopt whatever set of moral principles is expressed by your intuitive responses to different moral dilemmas'). We might, for example, ground it in a further *rational* principle of the form 'maximise your own good', provided that it was true, owing to certain facts and logical entailments, that living by the principles entailed by your intuitions was a key element of maximising that good. But then we have to ground *that* principle. So we might ask: Why should we maximise our own good? If we say that we should do so because it makes us happy, or content, or justified by our own lights, then we have hit *tautology*, but then what else is there to say?

The first thing to say is that for a principle to *be* a principle, it must be a principle *for* a certain kind of creature. We do not, for example, think sheep should live by the same principles as ordinary human adults, or that rocks should treat other rocks the way that they themselves would like to be treated. The second is that for a principle to be *justified*, it has to be grounded in facts about the kind of creature it is intended for. Perhaps these are facts about the way that creature thinks (all your normative thoughts *express* the categorical imperative), or about the way they behave (all your actions *express* a preference for liberal political principles), but either way, they are still *facts*. This means that there are no ultimate principles; only *ultimate facts*. It also means that all *imperatives* are, when properly understood, *hypothetical imperatives*, given that they depend upon contingent features of the motivational make-up of the creatures for whom they are intended (which might of course turn out to be perfectly altruistic, and thus perfectly 'moral', in at least one sense of the term); and that all *external reasons*, however abstract they might appear, ultimately depend on *internal reasons* derived from those contingent features (Williams, 1981, 101-113). But we do not need to muddy the conceptual waters here. What matters here is the simple truth, if I am right, that *principles*, in order to be justified for a particular kind of creature (e.g. a human being), have to be grounded in *facts about that creature*.

Now, that does still leave us with the question of the *particular* relevance of the behavioural trends normative behaviourism invokes, even if it is true, in general, that all principles ultimately depend on facts. How do we respond to that? In short, my claim is that such trends only *justify* a particular set of political principles provided that certain *facts* about human preferences render those trends relevant in the right way. There are two ways of revealing such preferences. First, we study our normative thoughts (mentalism). Second, we study our behaviour (normative behaviourism). In either case, we might just think that, if human beings *do* show an observable behavioural (or discoverable mental) preference for egalitarian-liberal-democracy, then that is *all we need* in order to justify the principles of that system. But is it really? The problem here is that, presumably, not *everybody* expresses the same preferences, in which case we have to ask how such principles could be justified to those who just *lack* the observable/discoverable preferences in question.

This is an important problem, and again there are two routes we might take, the first of which involves stressing, again, that it affects mentalism just as much as it affects normative behaviourism, given that, if an individual lacks either the right normative thoughts or the

motivation to act on them, we still have to work out what could be said to convince them (as opposed to coerce them) to cooperate. The second response, by contrast, turns upon the fact that, within the aggregated group of individuals lacking this preference, different individuals have *different reasons* for eventually accepting the principles in question.

Let's expand on that second point. Consider here that some people will eventually accept egalitarian-liberal-democratic principles because they realise that they will be safer, given the relatively low crime levels, than they would be in other systems (the *safety* motivation). Others will do so because they see that it is deemed more palatable than any other by the wider population (the *reasonableness* motivation). Still others will do so because they think that fewer people suffer under this system than under any other (the *compassion* motivation). As a result, there comes a point where we could say that, because these are the principles that convince the *highest* number of people, and because no other set of principles could do *more* than this in any relevant way, there is nothing more that political philosophy can say on the topic – assuming of course that there is *no more harmony to our normative thoughts than there is to the described actions*.

That assumption is crucial. Consider here that, up until now, there has been no need to pass judgement on mentalism, just so long as normative behaviourism either reinforces its conclusions, or helps us in areas where mentalism struggles. This lack of urgency has been important, given the lack of space for dealing properly with both these merits *and* the problem of viability discussed so far. Now though the contrast is important, given that there are good reasons to think that insurrection and crime provide a clearer and stronger picture of our political-principle-preferences than can be found in our normative thoughts: *clearer* because there is more consensus in these actions than there is in our normative thoughts, and *stronger* because those actions, by their very nature, require greater commitment than can be generated from the safety of our armchairs (Floyd, 2017b, 120-165).

Of these two points, the clarity issue is particularly important, given that, if our normative thoughts are too *inconsistent* for the purposes of conventional (and thus mentalist) political philosophy, then we need not worry about their relative *strength*, given that they would be unable to convincingly generate any *one* set of political principles, let alone one that could rival the principles generated by normative behaviourism. My premise here is that our normative thoughts just *are* that inconsistent. They are inconsistent in the sense that (1) there is extensive conflict, within each mind, and within each type of thought (e.g. our considered judgements clash with each other); (2) there is extensive conflict in each mind between different types (e.g. our considered judgements clash with our intuitive choices of abstract principle); (3) there is extensive conflict between the different normative thoughts of different individuals; and (4) this conflict is so extensive as to be beyond the rescue of any plausible form of reflective equilibrium.

Clearly, there is insufficient space to fully support that premise here (Floyd, 2017b, 99-165). I simply note it in order to fully explain the kind of argument required by *any* attempt to justify a particular choice of normative behavioural measures, and thus any more complete attempt to prove the defensibility of normative behaviourism as a whole, given that, for normative behaviourism in general to be viable, it must be true that at least some behavioural measures are normatively relevant, and not just in a general sense, but also in the particular sense of being able to trump, for example, our intuitions or considered judgements. Or, put differently, any such attempt would ultimately have to deal with the *relative merits* of

mentalism and normative behaviourism, given that even successful forms of normative behaviourism (in the sense of finding *very reliable* patterns of behaviour) have to be able to explain why, ultimately, we should prefer principles based on how we *do* behave to ones based on how we *think we should* behave.

### **1.8 Normative behaviourism defended, pt. 2: Whose behaviour? Which actions?**

We turn finally from defending normative behaviourism's general viability to defending the two types of behaviour it studies. In order to do so, I make a comparative case, and argue that insurrection and crime provide a *better* measure of political principles than three other possibilities – *happiness metrics*, *tax-evasion*, and *migration*<sup>16</sup>. This comparison draws on three problems that have not yet been fully illuminated: *domination*; *hot-headedness*; and *mixed results*<sup>17</sup>. The worry of the first is that a lack of insurrection and crime might simply be the result of people living in fear under an all-powerful regime. In this case, normative behaviourism rewards domination. The worry of the second is that those who are angry and violent are more likely to rebel or break the law. In this case, normative behaviourism rewards hot-headedness. The worry of the third is that we might encounter regimes in which crime points one way but insurrection another. In this case we face mixed results, and would not know what to reward.

Consider first then the possibility of using happiness metrics as a measure of political principles. Here we would not be tracking behaviour at all, but also not normative thoughts as mentalism does. Instead, we would judge political principles according to the happiness scores they generate in practice. Such a score might track the average level of surveyed happiness, or the happiness of the least happy quartile of the population, but either way, we would be matching principles to some kind of survey-data. This would be particularly tempting when considering the problem of *domination* just noted. Think here, for example, of authoritarian societies in which people are peaceful and law-abiding simply because they are dominated by an oppressive regime. A survey might cut through this and reveal the discontent below. Similarly, imagine a regime in which men rose up every time someone threatened to undermine their domination of 'their' women<sup>18</sup>. Again, surveys might reveal this by revealing the unhappiness of those whose outward behaviour is perfectly conformist.

There are, however, several problems with this approach. First, justification: Justifying any particular metric of happiness is going to require mentalist justification of the kind explained above, and if that fails for the reasons given, then we lack convincing reasons for tracking one metric over another, or indeed happiness versus something else, such as 'capabilities', or reported mental health problems. Second, gauging the strength of the feedback given: Teenagers, for example, are routinely miserable, and perhaps especially so in capitalist democracies where images of unattainable luxury are especially ubiquitous. However, despite such misery, they normally manage to live meaningful lives as they get older, whereas those who took part in the 'Arab Spring' made a different choice. In short, the deeds of the dominated speak louder than the desires of the decadent.

Nevertheless, how does normative behaviourism address the domination problem? As noted earlier when discussing the suffragettes/suffragists, normative behaviourism does so by recommending a particular system – egalitarian liberal democracy – that encourages both the dominated to protest *and* the non-dominated to help them. As a result, whether we suspect there are 'silent' dominated individuals around us or not, there is always an incentive to

conduct new political experiments to see if we can identify and/or help them, even if the full wisdom of those experiments arrives further down the line in terms of the insurrection and crime patterns eventually observed. Admittedly, this then raises the issue of how much normative behaviourism could have helped us *before* we had ‘experimented’ with democracy – an issue I put to one side for now - yet even then there would be some incentive to minimise domination as a form of inoculating against instability<sup>19</sup>. After all, rebellions are rarely in the interests of the actor being rebelled against, meaning that even King Arthur sometimes has to ‘play Kant’ (Williams, 2005, 66).

Consider second the idea of using tax-evasion as a measure of the different political systems, and thus principles, that encourage or discourage it. As a distinctive *act*, and one that conceivably expresses a verdict on the government trying to collect the tax, it has promise, and was precisely the kind of civil disobedience famously recommended by Thoreau, who withheld his own tax in protest against slavery and his country’s war with Mexico. In addition, it also avoids the problem of hot-headedness, as failing to complete a tax return avoids the risks of violence of crime and insurrection. As a result, for respectable, middle-class folk, as opposed to angry and impoverished young men, it is morally palatable in a way that insurrection, in particular, is not.

The question is whether it tracks what we want it to track. Consider, for example, the recent history of tax-evasion in Greece. Was it a protest against capitalism? Inequality? Excessive defence spending? Hardly. As with global tax-evasion more generally, it is largely a self-serving act, focussed on wealth-maximisation, and without regard for the regime one might undermine through lack of support. In fact, if one tried to measure political principles according to those systems that produce the most tax evasion, one would get not only a very confused picture, due to variations connected to enforcement, but also one that suggests that what people want above all else is a state stripped of just about everything that makes it a state.

Yet would that be a *mistake*? After all, many wealthy individuals *are* libertarians. The problem though is that yet another action carries even more weight – the choice of where these individuals *live*. Although their money travels to the Caribbean, their houses and families remain in more social-democratic locations. So: *some* actions speak louder than *others* – even if this does open the door for making migration a better measure than insurrection or crime. But we get to that in a moment. For now, the point is that tax-evasion fails to displace insurrection or crime. Yes, we might worry about a bias towards the naturally hot-headed, but bear in mind that when tracking actions that express *anger* at a certain set of institutions – and thus principles – we are revealing *serious* convictions. Actions that express only mild irritation at this or that tax bracket are no guide at all to political principles<sup>20</sup>. People always complain. The acid-test is when they move *themselves* to action, and not just their money.

Saying that tax-evasion is a worse measure of principles, however, hardly puts the problem of hot-headedness out of sight, because again, we still have this worry that some people might be more likely to rebel than others. Those who are naturally more violent perhaps, but also the young more than the old, and men more than women. As a result, we might want to try something new here, and start giving greater ‘weight’ to expressions of discontent from those from whom we least expect it – violent pensioners, or criminal women, for example<sup>21</sup>. Yet that, I think, would be a mistake. First, the fact that certain groups of people ‘traditionally’



rebel less might just reflect their ‘traditionally’ having been better off – the rich for example. We would not, I assume, want to give greater ‘weight’ to a rebellion by millionaires. Second, we should be wary of ascribing permanent tendencies to groups based on historical contingencies. Women might be less prone to violence *so far*, but will they always be so, even on the other side of patriarchy? As a result, I suggest a less radical, though more limited two-part solution than weighting. First, we note once more that normative behaviourism recommends egalitarian liberal democracy. Second, we note that, under such an order, women and the old, to take just two key groups, are if anything *more* likely to vote. This means that even if young men ‘naturally’ rebel more than other groups, that is both (a) true of all systems, for the purposes of comparison, and (b) doesn’t stop us recommending a system in which that tendency has no ‘pay-off’ for them as a group (and indeed, it may simply reduce employability and increase incarceration).

The only downside to this solution is that it limits normative behaviourism by highlighting, one last time, the worry that it would not have been much good to us *before* we had egalitarian liberal democracy. It is, after all, quite *hard* to play Kant at the Court of King Arthur, just as it’s hard to avoid the conclusions that young men have more say than anyone else in such environments, and that they would have caused a lot of violence had that changed. Nonetheless, it’s also true that, as it happens, things *did* change, with later systems proving better than old at reducing crime and insurrection. And, in turn, it’s also true that mentalist political philosophy might have faced an almost identical problem. If, for example, political principles are to be based on considered judgements (Gaus, 2011, 174), and if such judgements in the past pointed to slavery, sexism, witch-burning, crusades, and so on, then like Rawls you might just have to accept a circular connection between liberal democracy and good political philosophy, given that only under the former do you have the conditions for truly ‘considered’ judgements, meaning that liberal democracy, in effect, creates the conditions for its own vindication (Floyd, 2016b; Floyd, 2017a).

Consider finally then the possibility of measuring principles by migration patterns. Here, we might say, the idea is that good principles bring people *in* whilst bad ones send them *away*. This is a promising idea, when one thinks that we are basing our principles, not on the question of what people hypothetically choose, but on what people actually choose. In other words, we have much of the appeal of normative behaviourism as already discussed, but without, at least, the worry of hot-headedness already canvassed (though the worry of domination might return, given that dominated people struggle to move). In turn, we might also have a solution to the problem of mixed results. If China, for example, minimises democracy and liberal rights, but also violence against women, whilst India offers the reverse, then insurrection and crime might struggle to pick the two apart<sup>22</sup>. Assuming, therefore, that (1) the two countries produce similar levels of those two things, but also (2) neither regime is more or less supported by historical precedent, it might be that only migration could separate them. In that case, even if migration was only a tie-breaker, it would be a crucial one. Once we know whether people are more likely to move from China to India or *vice versa*, we would be able to rank the two sets of principles those regimes express.

There are, however, several problems with this idea. First, contemporary migration rates owe more to wealth disparities, cultural openness, and border controls, than the political principles in play - even if all three things, in the long run, are a function of just those principles we want to measure. Nonetheless, in the short term, one might move to California to join a technology company, or Saudi Arabia to join an engineering firm, for reasons entirely

connected to fleeting economic trends or natural resource wealth, even when this is at odds with the political convictions of those involved. ‘Brain drain’, in other words, often takes people well against the current of their *political* will.

A further problem concerns the second assumption above – that of the historical records being equal. If, for example, we take China to be an authoritarian regime, and consider such regimes indicted by their historical record in the way that, say, the Polity Report<sup>23</sup> argues, then perhaps we should not be distracted by what happens during a few decades of economic boom. What we are seeing now would be an exception, not the rule. Yet *what if* China carried on serenely for the next thousand years? What if, in the long run, it showed lower insurrection and crime, lower tax evasion, high immigration, and even a leading happiness score? In that case, we should be clear, normative behaviourism has to change its conclusions. Its core argument is for a method of measuring principles, not any particular system, whether that be Chinese one-party rule or indeed some new system that has yet to be tried, but which one day someone experiments with.

Nevertheless, with this in mind, we should stress one last point here, namely that acknowledging this possibility also means keeping the domination problem firmly in mind, given that we need to be absolutely sure what the real basis is for ‘sustained’ low crime and insurrection. Is it contentment or is it just a function of new and incredibly successful surveillance technologies? So far, I have assumed that such technologies would crack in the end, whilst also stressing that, for example, a dictatorial regime in which the system had not yet cracked would still be indicted by analogy with all those in which it did eventually happen – just as a democracy would be indicted if all but one democracies had collapsed. But still, what if a non-democratic regime of some kind just kept going? How long is *long enough* to change the verdict of normative behaviourism? A few years, clearly, is too little, whilst a few thousand seems an arbitrarily high threshold in order to ‘fix’ the results. And again, how are we to know whether that stems from high satisfaction, or just a drone in every home, a microchip in every neck, and as Orwell put it, a boot stamping on a human face - forever<sup>24</sup>? What does the data really tell us then?

As things stand, I leave this problem open. One option, clearly, is still just to look, comparatively, at the record of different regime types, and hope that the ‘truth’ will out. Another would be to admit some kind of tie-breaking criteria beyond insurrection and crime, such as suicide rates, online protests, or perhaps still migration, but this time maybe *unskilled emigration* rather than *skilled immigration*. Think here, for example, of the North Korean soldier who, in 2017, managed to cross the de-militarized zone, peppered with bullets and ridden with worms. Here we might say, to coin a pair of phrases, that ‘pain drain’ tells us more than ‘gain drain’, given that unskilled labour only moves when it is desperate, given the difficulty of leaving, the risks of travel, and the likely reception awaiting it when it arrives elsewhere. So, rather than judging regimes ‘good’ when they attract ‘the brightest and the best’, we might instead judge them bad when the worst-off start to leave. For now, to be clear, I am not convinced by these options, but all the same, when faced with regimes that display long-term stability for unclear reasons, these might be experiments ‘within’ normative behaviourism that are worth undertaking.

## References

ACLP Dataset: <http://politics.as.nyu.edu/object/przeworskilinks.html>

- Cohen, G.A. (2008). *Rescuing Justice and Equality* (Cambridge, MA: Harvard University Press).
- Floyd, J. (2010). Should political philosophy be more realistic?, *Res Publica*, 16(3), 337-347.
- Floyd, J. (2011a). From historical contextualism, to mentalism, to behaviourism. In J. Floyd & M. Stears (Eds.), *Political Philosophy versus History? Contextualism and Real Politics in Contemporary Political Thought* (pp. 38-64). Cambridge: Cambridge University Press.
- Floyd, J. (2011b). Relative Value and assorted historical lessons. In J. Floyd & M. Stears (Eds.), *Political Philosophy versus History? Contextualism and Real Politics in Contemporary Political Thought* (pp. 206-225). Cambridge: Cambridge University Press.
- Floyd, J. (2016a). Normative behaviourism and global political principles. *Journal of International Political Theory*, 12(2), 152-168.
- Floyd, J. (2016b). Analytics and Continentals: Divided by nature but united by praxis? *European Journal of Political Theory*, 15(2), 155-171.
- Floyd, J. (2017a). Rawls' methodological blueprint. *European Journal of Political Theory* 16(3), 367-381.
- Floyd, J. (2017b). *Is Political Philosophy Impossible? Thoughts and Behaviour in Normative Political Theory*. Cambridge: Cambridge University Press.
- Freedom House: <http://www.freedomhouse.org>.
- Galston, W. (2010). Realism in Political Theory. *European Journal of Political Theory*, 9(4), 385-411.
- Gaus, G. (2011) *The Order of Public Reason*. Cambridge: Cambridge University Press.
- Gilabert, P. & Lawford-Smith, H. (2012). Political Feasibility: A Conceptual Exploration, *Political Studies* 60(4), 809-825.
- Geuss, R. (2009). *Philosophy and Real Politics*. Princeton, NJ: Princeton University Press.
- Gutmann, A. & Thompson, D. (2004). *Why Deliberative Democracy?* Princeton, NJ: Princeton University Press.
- Hamlin, A. & Stemplowska, Z. (2012). Theory, Ideal Theory and the Theory of Ideals. *Political Studies Review*, 10(1), 48-62.
- Horton, J. (2010). Realism, liberal moralism and a political theory of modus vivendi. *European Journal of Political Theory*, 9(4), 431-448
- Jubb, R. (2012). Tragedies of non-ideal theory. *European Journal of Political Theory*, 11(3), 229-246.
- Kelly, M. (2000). Inequality and Crime. *The Review of Economics and Statistics*, 82(4), 530-539.
- Lawford-Smith, H. (2012). Understanding Political Feasibility. *The Journal of Political Philosophy*, 21(3), 243-259.

- Mason, A. (2010). Rawlsian Theory and the Circumstances of Politics. *Political Theory*, 38(5), 658-683.
- Mills, C.W. (2005). "Ideal Theory" as Ideology. *Hypatia*, 20(3), 165-184.
- Newey, G. (2010). Two dogmas of liberalism. *European Journal of Political Theory*, 9(4), 449-465.
- Neumayer, E. (2005). Inequality and violent crime: evidence from data on robbery and violent theft. *Journal of Peace Research*, 42(1), 101-112.
- Okin, S.M. (1989). *Justice, gender, and the family*. New York, NY: Basic Books.
- O'Neill, O. (1987). Abstraction, idealization, and ideology in ethics. In Evans, J.D.G. (Ed.), *Moral philosophy and contemporary problems* (pp. 55-69). Cambridge: Cambridge University Press.
- Orwell, G (2000). *Nineteen Eighty-Four*. London: Penguin.
- Pankhurst, E. (1979). *My Own Story*. London: Virago.
- Polity IV Dataset, <http://www.systemicpeace.org/polity/polity4.htm>
- Przeworski, A., Alvarez, M.E., Cheibub, J.A., Limongi, F. (2000). *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. (1999). *Collected Papers*. Cambridge, MA: Harvard University Press.
- Robeyns, I. (2008). Ideal Theory in Theory and Practice. *Social Theory and Practice*, 34(3), 358-360.
- Runciman, D. (2012). What is Realistic Political Philosophy? *Metaphilosophy*, 43(1-2), 66-67.
- Sabl, A. (2002). When bad things happen from good people (and vice-versa): Hume's political ethics of revolution. *Polity*, 35(1), esp. 73-92.
- Scanlon, T.M. (1998). *What we Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schmitz, D. (2011). Nonideal Theory: What It Is and What It Needs to Be, *Ethics*, 121(4), 772-796.
- Sen, A. (2009). *The Idea of Justice*. London: Allen Lane.
- Simmons, A.J. (2010) Ideal and Nonideal Theory, *Philosophy and Public Affairs*, 38(1), 5-36.
- Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(3), 229-243.
- Valentini, L. (2009). On the Apparent Paradox of Ideal Theory. *The Journal of Political Philosophy*, 17(3), 332-355
- Waldron, J. (1999). *Law and Disagreement*. Oxford: Oxford University Press.
- Wiens, D. (2011). Prescribing Institutions Without Ideal Theory. *The Journal of Political Philosophy*, 20(1), 45-70.

Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.

Williams, B. (2005). *In the Beginning was the Deed: Realism and Moralism in Political Argument*. Princeton, NJ: Princeton University Press.

Wilkinson R. & Pickett, K. (2009). *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Allen Lane.

Ypi, L. (2010). On the Confusion between Ideal and Non-ideal in Recent Debates on Global Justice. *Political Studies*, 58(3), 536-555.

---

<sup>1</sup> I first proposed normative behaviourism in (Floyd, 2011a). Its most comprehensive statement is (Floyd, 2017b). For a recent application to international politics, see (Floyd, 2016a).

<sup>2</sup> These thoughts include what we tend to call ‘moral’ thoughts, but are not restricted to them, given that for many scholars, from Hobbes to Gauthier, the justification of political principles is at least as much a matter of rationality as it is of morality – a point of particular relevance here, given the tendency of realists to focus on ‘order’ over ‘justice’, e.g. (Williams, 2005).

<sup>3</sup> I believe this list is comprehensive, but it is enough here if it is illustrative.

<sup>4</sup> The most important caveat is the *coherentist* one, normally presented in the form of *reflective equilibrium*, according to which ‘rogue’ normative thoughts, i.e. thoughts that contradict most of our other thoughts of this kind, should be rejected as unreliable, and thus not genuine *expressions of*, and therefore evidential *grounds for*, the underlying principles we are after.

<sup>5</sup> The *Polity* dataset can be found at <http://www.systemicpeace.org/polity/polity4.htm>. The *ACLP* (Alvarez, Cheibub, Limongi, Przeworski) dataset was the basis of (Przeworski, Alvarez, Cheibub, Limongi, 2000). The original dataset can be found at <http://politics.as.nyu.edu/object/przeworskilinks.html>. The *Freedom House* dataset can be found at <http://www.freedomhouse.org>.

<sup>6</sup> I should though note a possibility that I will not explore in this article: that normative behaviourism might be impossible at the level I suggest here, though still possible at a more general level. It might, for example, be true that certain kinds of behaviour can function as grounds for certain kinds of principle, even if the particular kinds of behaviour I describe cannot function as grounds for political principles in the way that I suggest. In that case, although it might still be true actions speak louder than thoughts in a certain way, it would not be true that normative behaviourism contributes to political philosophy in the way that I imagine it does.

<sup>7</sup> Both authors take their lead from (O’Neill, 1987).

<sup>8</sup> Mills takes his lead here from (Okin, 1989). For helpful discussion of both this example and the problem of ‘bad idealisations’, see (Robeyns, 2008).

<sup>9</sup> For the importance of ‘experimentation’ to normative behaviourism, see (Floyd, 2017b, 244-249).

<sup>10</sup> Admittedly, this does raise two important questions, urged on me by an anonymous reviewer for this journal (R3): (1) how radical are experiments ‘allowed’ to be within normative behaviourism?; and (2) Could normative behaviourism have guided us before there was serious experimentation with liberal democracy? I cannot adequately address these questions here, though touch on (2) in the final sub-section of this article. For more on this front, see again (Floyd, 2017b, 244-249). For an excellent discussion of how hard it is to judge the wisdom of political experiments without historical precedent to guide us, see (Sabl, 2002, 78-32).

<sup>11</sup> For an overview of the kinds of reasons realists invoke when supporting one or the other conclusion, see (Galston, 2010).

<sup>12</sup> According to the data on which normative behaviourism relies, these are systems that practice a particularly fair and deliberative kind of democracy against an optimally liberal-egalitarian backdrop. The ‘across all values’ clause is important because comparative/transitional approaches have to be balanced not just for one value but across all the various procedures and states of affairs that human beings find valuable.

<sup>13</sup> The pioneering deliberative variant is (Gutmann & Thompson, 2004). The majoritarian equivalent is (Waldron, 1999).

<sup>14</sup> Both Horton and Newey (see fn.19) acknowledge this difficulty. For further discussion, see (Runciman, 2012).

<sup>15</sup> For the realist version, see (Galston, 2010, 392-394). For the non-ideal theory version, see (Ypi, 2010) and (Wiens, 2011).

---

<sup>16</sup> The suggestion of happiness was helpfully made by one anonymous reviewer for this journal (R1); the suggestions of migration and tax evasion were helpfully made by another (R2).

<sup>17</sup> I am also grateful to R1 and R2 for highlighting these three problems.

<sup>18</sup> I owe this example to R2.

<sup>19</sup> See fn. 10.

<sup>20</sup> There's a parallel here to Rawls' view that minor changes to tax codes were insufficient grounds for civil disobedience. See (Rawls, 1971, 363-391).

<sup>21</sup> This clever suggestion was put to me by R3.

<sup>22</sup> This contrast was suggested to me by R2.

<sup>23</sup> See fn. 7.

<sup>24</sup> I thank R3 for pointing out the challenge posed by societies like 'Airstrip One'. See (Orwell, 2000, ch. 3).